An interactive graphics system for comparing and model building of macromolecules

Brandenburg, N.P.; Dempsey, S.; Dijkstra, B.W.; Lijk, L.J.; Hol, W.G.J.

Published in:
Journal of Applied Crystallography

DOI:
10.1107/S0021889881009357

Publication date:
1981

J. Appl. Cryst. (1981). **14**, 274–279

An Interactive Graphics System for Comparing and Model Building of Macromolecules

By N. P. Brandenburg

Laboratory of Chemical Physics, University of Groningen, Nijenborgh 16, *Groningen, The Netherlands*

S. Dempsey

Chemistry Department, B-014, *University of California at San Diego, La Jolla, CA* 92093, *USA*

and B. W. Dijkstra, L. J. Lijk and W. G. J. Hol*

Laboratory of Chemical Physics, University of Groningen, Nijenborgh 16, *Groningen, The Netherlands*

(*Received* 8 *December* 1980; *accepted* 9 *February* 1981)

Abstract

An interactive graphics program is described for use with the Evans & Sutherland Picture System 2 which is suitable for the comparison and refinement of protein structures. Several protein molecules and electron density maps can be viewed simultaneously, while great flexibility exists in creating, modifying and manipulating the picture on the screen. As the program is file oriented, it can be run on a small computer system with only 32 K memory.

Introduction

The benefits of computer graphics in the field of protein crystallography have been recognized for quite some time, and in the past decade several modeling programs for macromolecules were developed (Barry & North, 1971; Langridge, 1974; Morimoto & Meyer, 1976). In addition to software for these purposes, special hardware has also been designed. For instance, in the case of the molecular MMS system, special emphasis was placed on the advantage of hardware components dedicated to macromolecular modeling purposes (Barry, Bosshard, Ellis & Marshall, 1974). In the North Carolina system, a large variety of interactive devices has been built and used with considerable success in protein model building (Wright, 1972; Tsernoglou, Petsko, McQueen & Hermans, 1977).

Nowadays a small number of interactive graphics systems equipped with a powerful special-purpose microprocessor are commercially available. For the Vector General 3400 a model-building program has been written (Jones, 1978). Another modern graphics system is the Evans & Sutherland Picture System. Programs for the building and manipulation of macromolecules on this system have been developed by Dempsey (1977) and by Diamond (1978). For a newer version of this Evans & Sutherland product, the Picture System 2, the programs of Jones have been adapted (Jones, 1980).

We wish to describe in this communication a program *GUIDE* (Groningen University Interactive Display of Enzymes) for use with the Picture System 2. This program has been developed with two main applications in mind:

(i) the improvement of the fit of a molecular model into an electron density map;

(ii) the comparison of proteins and other macromolecules which are related in structure and/or function.

Basic philosophy

Before going on to describe the features of *GUIDE* in some detail it may be useful to touch briefly upon the basic ideas of the program.

It was felt imperative that a program like *GUIDE* should have the possibility of studying several models simultaneously. As the number of known protein structures is growing rapidly, this feature allows for the comparison of active sites and folding patterns of different protein structures. Alternatively, one should be able to show one electron density 'model' together with several macromolecular models in order to investigate, for example, the best fit of a number of conformations to an electron density map. Another useful option is the possibility of comparing several electron density maps in order to study, for example, the progress made in subsequent refinement steps.

In the most general mode, then, more than one electron density map may be shown together with more than one molecular model. This may be desired for studying the mode of binding of small molecules to macromolecules.

It was furthermore considered essential that the program be very flexible in the creation of the picture. No limitations exist as to which fragments of a macromolecule can be selected for appearance at the screen. The user is able to compose the picture from any number of fragments of the molecule(s) under investiga-

* To whom correspondence should be addressed.

0021-8898/81/040274-06$01.00

tion. This is obviously of great value in comparative studies of active-site geometries or folding patterns. But also, when fitting models into electron density maps it is essential to see the surroundings of the part presently being manipulated as, for example, hydrogen-bonding patterns may be of great help in finding the best fit. In addition, one can choose to focus on the backbone of a protein molecule by creating a picture in which only the C^{α} atoms are selected for display on the screen.

For model-fitting purposes, one should obviously have the possibilities of global rotations, translations and zooming. Further requirements are movement of model *versus* electron density, rotations about bonds, *etc.* In addition, it proved to be important to have a way of disconnecting one part of a molecule from the rest of the molecule. Such a disconnected part is called a 'molten zone'. Several molten zones can be defined simultaneously and may even be nested if so desired. The molten-zone options are important during fitting procedures. They allow the user to disconnect a range of residues, or a single phenyl ring, or a peptide group from the rest of the protein molecule and rotate and translate it freely in space. After an optimal fit has been obtained the molten zone can be reconnected to the macromolecule while the correct geometry can be restored in a subsequent regularization step. This regularization occurs at present on the CDC of the University of Groningen Computer Centre.

As to the man–machine interface, it was considered important to limit the use of the keyboard as input medium as much as possible. Instead, interactive devices such as, for example, a data tablet are employed for editing functions and dynamic alterations of the picture. This allows the user to concentrate continuously on the objects at the screen without having to divert his attention to the keyboard. In addition, all tablet commands are on a single menu eliminating the need for paging through several menu's.

As to the way of incurring depth vision, the philosophy has been that the user should be able to choose from several options. Depth cueing is provided and repositioning the 'clipping planes' allows control of the extent of depth cueing by the user. An automatic rocking mode, with variable rocking speed, is provided while others may prefer an automatic continuous rotation of the picture. In addition, side-by-side stereoscopic pairs can be shown which can be viewed by a mirror system as developed at the Biozentrum in Basel (Professor J. N. Jansonius, personal communication).

For convenience in alterations of functions and additions of options, it was considered important that the program be written in a high-level language (Fortran in this case) and have a very well defined structure. The latter has been achieved by the use of a large number of subroutines which also allows easy definitions of overlay structures. It was furthermore thought to be important to limit the number of files as much as possible. This would make the use of the system easy. The system uses at present only three different kinds of files: the molecular-model files, the density-contour files and one file which provides atom and residue name information. The first two files can be created very easily on the CDC of the University of Groningen Computer Centre and are subsequently transported to the display system.

Computer configuration and general data flow

A major part of our graphics system is the Evans & Sutherland Picture System 2, which comprises a powerful microprocessor (the 'picture processor') and a high-resolution scope. The system is equipped with 48 K refresh memory and the following interactive devices: 16 control dials, 16 function switches and a high-resolution Summagraphics data tablet. The Picture System is interfaced to a Digital Equipment Corporation PDP-11/34 minicomputer with 32 K of MOS memory. The input/output devices comprise several RK05 disks, a DEC writer and a 9600 baud line to the CDC.

In what follows, the general data flow is described as going from the University's CDC machine to the screen of the display system. It may be pointed out, however, that molecular models after modification by *GUIDE* are also sent back from the PDP-11/34 to the main computer.

The coordinate (and contour) data reside on the CDC computer. After being processed there they are sent to the PDP-11/34 and stored on disk. *GUIDE* works with a copy of this data base and, according to the editing commands given, modifies the status bits of the atoms in the database. After completion of the editing stage, a so-called 'linear display list' is created, with subroutines from the Evans & Sutherland software package. This list is a collection of picture-processor commands and data which are accessed on a DMA basis by the special-purpose microprocessor. In this way full advantage is taken of the speed of the microprocessor as the PDP-11/34 does not need to be interrupted for each data transfer step. *GUIDE* controls the dynamic motion of the picture by polling the control dials and constant recalculation of the orientation matrices which are either placed in front of or inserted into the linear display list.

After processsing the linear display list, the data are stored in the refresh buffer. The refresh buffer is used in the double buffer mode, *i.e.* it is divided into two parts. From one part the picture on the screen is refreshed with a refresh rate of 50 Hz, while the second part is being filled with newly processed data. At appropriate intervals the functions of these two halves are switched. This process uncouples the 'refresh rate' of 50 Hz from the 'update rate' which may be 10 Hz without causing flicker of the picture on the screen.

Data bases

GUIDE can handle two kinds of disk-resident data

bases for the display of molecular models and electron densities, respectively. The data base for molecular models is kept on a random-access file: most records are atom records and contain coordinates, atom and residue type codes and status information. Status bits are used by the program to decide, amongst other things, in which mode (solid or dashed lines; with or without labels, *etc.*) the atoms are to be drawn on the screen.

Besides coordinate, name and status information for each atom, the connectivity of the atoms has to be known. There are several ways to provide the connectivity information. It is possible to append this information explicitly to each atom record in a data base. Another solution is to calculate distances between atoms, suitable distance criteria then lead to the correct connectivity (Jones, 1978). Since the first solution leads to a relatively large data base, and since the second may lead to difficulties in the case of crude coordinates obtained, for example, from a wire model, we have adopted the principle of implicit connectivity. The atom records are interspersed in such a way with 'dummy' atoms that connecting each atom with the next atom (real or dummy) results in the desired picture of the molecule.

The data base for the display of electron density models (chicken-wire representation) is in fact a large set of contour vectors, prepared on the CDC 170 from an electron density distribution. For each individual contour, being a closed loop or ending at the edges of the electron density block, the maximum and minimum coordinate values in x, y and z are stored in the data base. These data are, amongst others, used by *GUIDE* to select automatically the desired contours corresponding with a range of atoms visible on the screen. This approach is different from the solution to divide the contour data into a set of small 'building blocks' (Diamond, 1978). The latter approach will lead, generally, to a quick assembly of large contour regions, at the expense, however, of directory overhead.

Features

Modifications of the picture can be divided into 'editing functions' and 'dynamical update operations'. For these types of functions, four types of interactive devices are available: (i) DEC writer (ii) function switches (iii) data tablet and (iv) control dials. The DEC writer is only used for entering names of files to be fetched or created, for error messages and for keeping a record of the editing commands given. The function switches are used for a limited number of well-defined on/off operations such as switching to 'stereopair set mode' and switching to the mode for entering new data bases. All the remaining editing functions are performed by means of repositioning the pen on the data tablet, while all dynamical update operations are controlled by the dials.

The set of edit commands which can be invoked by tablet operations is given in Table 1, together with a brief explanation of each command. These commands allow the selection of the residues of a model, cause dials to govern rotations about certain bonds, identify a

Table 1. *Editing functions*

Command	Function
BAC+	Show backbone, *i.e.* connect C^{α} atoms
BAC−	Delete backbone
BLK+	Blink
BLK−	Stop blinking
CNT	Show contours
DEL	Delete atoms
DIS	Add atoms to ones present on the screen
GT1	Specifies first atom of a distance
GT2	Specifies second atom of a distance
LBL+	Display atom/residue labels
LBL−	Remove labels
MLT	Install molten zone
ORG	Define origin
RIG+	Set rigid bits, *i.e.* no rotation about bond allowed
RIG−	Clear rigid bits, *i.e.* rotation about bond allowed
ROT	Install rotation about a bond
SHO	Show atoms after deleting existing ones
TRN	Install atom to be moved individually
TXT+	Draw lines dashed
TXT−	Draw dashed lines now solid

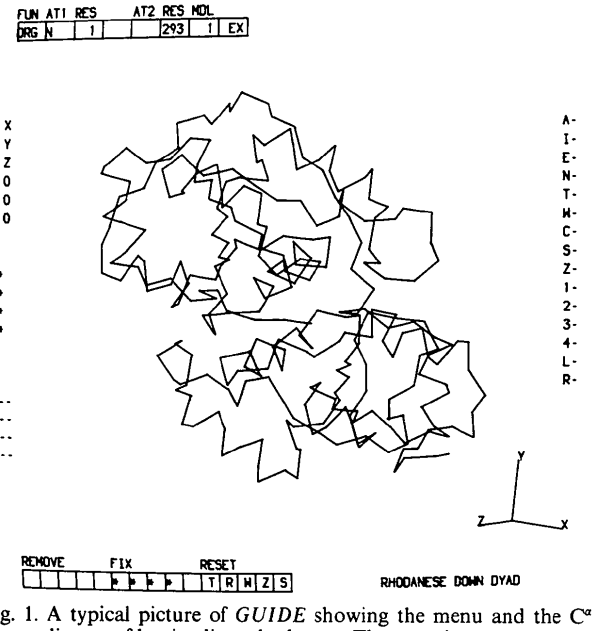

Fig. 1. A typical picture of *GUIDE* showing the menu and the C^{α} coordinates of bovine liver rhodanese. The most important menu areas are the edit commands in the upper left corner and the remove, 'fix' and reset commands in the lower left line. By moving the tablet pen downwards along a vertical line from the FUN, ATI, RES, *etc.* positions the function, atom name and residue number displayed is constantly changed. This allows easy selection of the function to be executed and of the atom (or atom range) to which the function will be applied. The function will be executed upon touching the EX area. The lower left line allows resetting, removal or 'fixing' of rotations about bonds, of molten zones, of individual atoms, of global rotations, translations, slab size, *etc.*

'flying atom' which subsequently can be moved by dials independently of the other atoms, specify atoms to be labeled, define 'molten zones', *etc.* The majority of these functions can be applied to any range of atoms, from a single atom up to all atoms in a model. In addition, residue types may be specified rather than residue numbers, if so desired. This allows the user to show all arginines of a model with a single command, for instance. These editing commands appear in the left upper corner of the screen as shown in Fig. 1. Other areas of the permanently displayed menu shown in this figure allow (i) fast activation and inactivation of the molten zones, (ii) resetting, fixing and removing rotations about bonds, (iii) invocation of automatic rocking or rotation modes, (iv) enlargement of the character size of the labels, (v) increase in the brightness of a particular model, (vi) removal of the 'distance reporting' between two atoms, *etc.* Simple tablet operations allow the user to select one, two or three contour levels to be shown in a 'density model' and to specify whether these levels should be drawn with solid or dashed lines.

Table 2. *Functions of control dials*

Dial number	Function
1	Rotation about bond number 1
2	Rotation about bond number 2
3	Rotation about bond number 3
4	Rotation about bond number 4
5	Rotation about bond number 5
6	Individual translation x
7	Individual translation y
8	Individual translation z
9	Zooming
10*	Global rotation x
11*	Global rotation y
12*	Global rotation z
13†	'Slab size', *i.e.* front and back clipping planes both move with respect to the origin of the picture.
14*	Global translation x
15*	Global translation y
16*	Global translation z

* Dials 10–12 and 14–16 affect a molten zone, and not the overall picture, if a molten zone is activated.

† Dial 13 controls the rocking velocity if this mode is initiated by the tablet ($W+$ in Fig. 1).

The control dials govern the dynamic operations of the picture as explained in Table 2. These include overall rotations and translations of the picture. When a molten zone is activated by means of a tablet pen action, the dials which previously governed these overall functions now control the movement of the active molten zone only. Subsequent pen movements allow a quick activation of another molten zone, or return to the 'global' mode.

Applications

The interactive graphics system described in this article has been used in the refinement of the structure of bovine liver rhodanese. This is a sulfur-transferring enzyme which functions *via* a two-step mechanism during which a stable rhodanese–sulfur intermediate is formed (Westley, 1973). The structure of this intermediate was initially determined at a resolution of 2·5 Å (Ploegman, Drent, Kalk & Hol, 1978) by the method of multiple isomorphous replacement. Fig. 2 shows the backbone of this protein molecule rotated in such a way that the similarity in folding between its two halves is obvious. This is a striking observation as the amino-

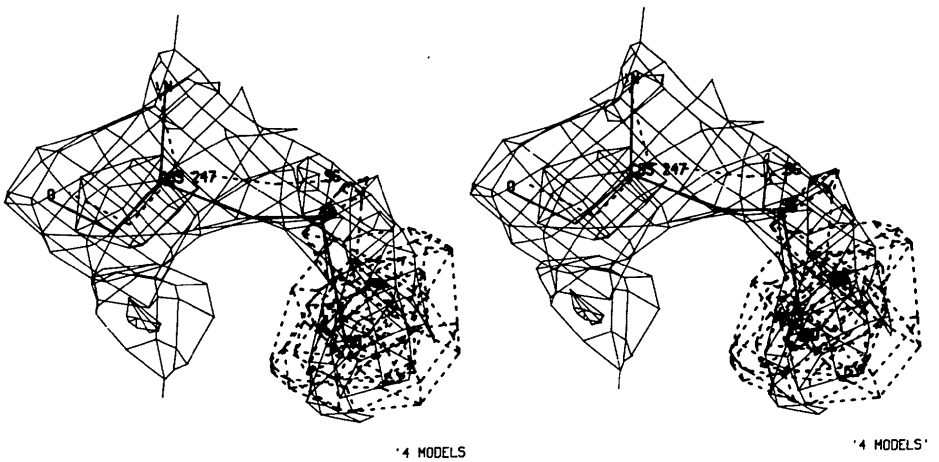

Fig. 2. A stereopicture showing the essential cysteine residue 247 of the bovine liver rhodanese–sulfur complex. The model in solid lines represents the coordinates, after several cycles of real-space refinement, of Cys 247 plus the extra sulfur atom which is bonded to the S^{γ} atom of this residue. The model with dashed lines is the initial model used for the refinement obtained after regularization of the wire-model coordinates. The solid contours represent the $(2F_o - F_c)$ electron density map calculated after several real-space refinement steps. The dashed contours show the difference electron density obtained by treatment of sulfur–rhodanese crystals with cyanide and illustrate clearly the removal of the extra sulfur atom by this procedure.

acid sequences of the two halves are very dissimilar indeed (Ploegman *et al.*, 1978).

After regularization of the wire-model coordinates the reliability factor of F_c *versus* F_o was 43%. The resolution has been extended to 2·2 Å while the refinement, with Diamond's real-space refinement procedure (Diamond, 1971), is presently at an intermediate stage with a reliability factor of 30%. Fig. 2 gives a good idea of a unique capability of *GUIDE* – it shows the initial regularized wire-model coordinates together with the latest refined model in a $(2F_o - F_c)$ electron density around residue 247 plus a difference electron density which shows that the extra sulfur atom can be removed from the rhodanese–sulfur complex (Ploegman, Drent, Kalk & Hol, 1979).

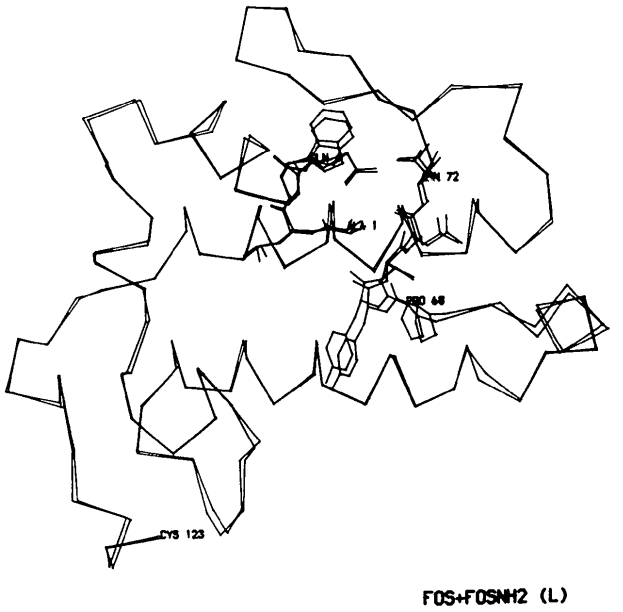

Fig. 3. The structures of native and transaminated phospholipase A_2: superposition of the C^α coordinates and all coordinates of a number of selected residues. These residues exhibit the largest changes in conformation when comparing the two structures. The figure illustrates the possibility of comparing general folding patterns simultaneously with detailed side-chain superpositions.

The graphics system has also been employed for the comparison of the structures of two phospholipase A_2 molecules. Phospholipase A_2 specifically catalyzes the hydrolysis of the 2-acyl linkage of phosphoglycerides (Van Deenen & De Haas, 1964). The pancreatic enzyme shows a tremendous increase of activity when the substrate concentration passes the critical micelle concentration (Pieterson, Vidal, Volwerk & De Haas, 1974). It was concluded that pancreatic phospholipase A_2 possesses, in addition to a catalytically active center, a so-called 'interface recognition site' which interacts with ordered substrates like micelles.

Interestingly, a minor modification (transamination of the N-terminal alanine residue) abolishes the tremendous increase of activity of the enzyme *versus* micelles while the activity towards monomers is virtually the same (Slotboom, 1981). Fig. 4 shows the structures of native and transaminated phospholipase A_2 after optimal superposition of the backbone atoms. Those residues which show the largest differences in structure are shown in full while the remaining parts of the molecules are represented by the C^α atoms only. It is obvious from this picture that the differences between

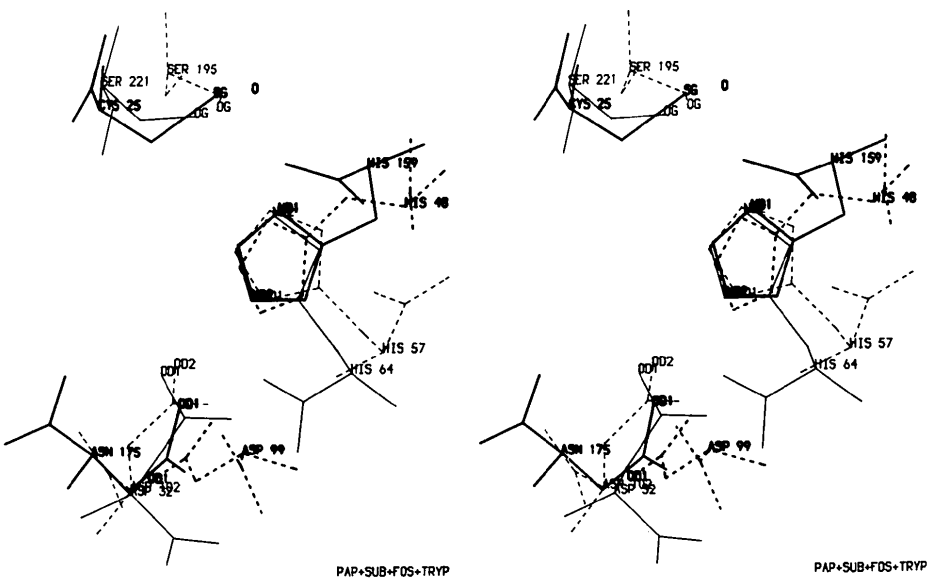

Fig. 4. Superposition of three essential active-site residues in four structurally unrelated hydrolases. The bright (= dark) solid lines correspond to the sulfhydryl protease papain, the light solid lines represent the serine protease subtilisin, the dark dashed lines (including the single 'O' of water 149 in the middle upper part of the figure) is bovine phospholipase A_2 and the light dashed lines show trypsin. The superposition was done manually on the display system with the papain residues kept fixed.

the two related enzymes are very small indeed, which makes the intriguing observation of the enhanced activity of the native enzyme towards micelles even more fascinating.

Attention has been drawn (Drenth, 1980; Drenth, Enzing, Kalk & Vessies, 1976) to the fact that the various classes of proteases utilize different groups in order to perform the same catalytic functions. It appears that this similarity can be extended to include pancreatic phospholipase. Fig. 4 shows a superposition of the active sites of the serine proteases subtilisin and trypsin, the sulfhydryl protease papain and phospholipase. The good superposition of the Asp–His–Ser triad in subtilisin and trypsin with the Asn–His–Cys triad in papain and an Asp–His–water triad in phospholipase A_2 is immediately obvious. It shows the strength of interactive display systems as described in this article for increasing our basic understanding of the structure–function relationship in protein molecules.

Discussion

Although the program allows a large number of operations to be performed, it is obvious from the above description that possibilities for improvement and expansion exist. The speed of the system should be improved considerably when an extra 64 K memory is used in the near future. Further improvements will result from the addition of a floating-point processor and the use of more advanced Evans & Sutherland software which will allow storage of the linear display list in the refresh buffer memory.

We wish to acknowledge the help of Mr L. Halie in setting up a library of protein molecule files for *GUIDE* from which Fig. 4 could be prepared conveniently. We furthermore would like to acknowledge the stimulating cooperation with the staff of the Groningen University Computer Centre.

This project was supported by The Netherlands Foundation for Chemical Research (SON) with financial aid from the Netherlands Organization for the Advancement of Pure Research (ZWO). Work at UCSD was supported by grant no. RR-00757 from the Division of Research Sources, National Institute of Health.

References

BARRY, C. D., BOSSHARD, H. E., ELLIS, R. A. & MARSHALL, G. R. (1974). *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **33**, 2368–2372.

BARRY, C. D. & NORTH, A. C. T. (1971). *Cold Spring Harbor Symp. Quant. Biol.* **36**, 577–584.

DEMPSEY, S. (1977). Personal communication.

DIAMOND, R. (1971). *Acta Cryst.* A**27**, 435–452.

DIAMOND, R. (1978). *BILDER* User's Guide.

DRENTH, J. (1980). *Recl Trav. Chim. Pays-Bas*, **99**, 185–190.

DRENTH, J., ENZING, C. M., KALK, K. H. & VESSIES, J. C. A. (1976). *Nature (London)*, **264**, 373–377.

JONES, T. A. (1978). *J. Appl. Cryst.* **11**, 268–272.

JONES, T. A. (1980). Personal communication.

LANGRIDGE, R. (1974). *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **33**, 2332–2335.

MORIMOTO, C. N. & MEYER, E. F. (1976). In *Crystallographic Computing Techniques*, edited by F. R. AHMED, pp. 488–496. Copenhagen: Munksgaard.

PIETERSON, W. A., VIDAL, J. C., VOLWERK, J. J. & DE HAAS, G. H. (1974). *Biochemistry*, **13**, 1455–1460.

PLOEGMAN, J. H., DRENT, G., KALK, K. H. & HOL, W. G. J. (1978). *J. Mol. Biol.* **123**, 557–594.

PLOEGMAN, J. H., DRENT, G., KALK, K. H. & HOL, W. G. J. (1979). *J. Mol. Biol.* **127**, 149–162.

SLOTBOOM, A. (1981). In preparation.

TSERNOGLOU, D., PETSKO, G. A., MCQUEEN, J. E. & HERMANS, J. (1977). *Science*, **197**, 1378–1381.

VAN DEENEN, L. L. M. & DE HAAS, G. H. (1964). *Adv. Lipid Res.* **2**, 167–234.

WESTLEY, J. (1973). *Adv. Enzymol. Relat. Subj. Biochem.* **39**, 327–368.

WRIGHT, W. V. (1972). Thesis, Univ. of North Carolina, Chapel Hill.